

FOCUS: INDUSTRIAL PROTEOMICS

In Vivo Labeling: A Glimpse of the Dynamic Proteome and Additional Constraints for Protein Identification

Rachel R. Ogorzalek Loo and Joseph A. Loo*

Department of Biological Chemistry, University of California-Los Angeles, Los Angeles, California, USA

Ping Du

Pfizer Global Research and Development, Groton, Connecticut, USA

Tod Holler

Pfizer Global Research and Development, Ann Arbor, Michigan, USA

Identities ascribed to the intact protein ions detected in MALDI-MS of whole bacterial cells or from other complex mixtures are often ambiguous. Isolation of candidate proteins can establish that they are of correct molecular mass and sufficiently abundant, but by itself is not definitive. An in vivo labeling strategy replacing methionine with selenomethionine has been employed to deliver an additional constraint for protein identification, i.e., number of methionine residues, derived from the shift in mass of labeled versus unlabeled proteins. By stressing a culture and simultaneously labeling, it was possible to specifically image the cells' response to the perturbation. Because labeled protein is only synthesized *after* application of the stress, it provides a means to view dynamic changes in the cellular proteome. These methods have been applied to identify a 15,879 Da protein ion from *E. coli* that was induced by an antibacterial agent with an unknown mechanism of action as *SpY*, a stress protein produced abundantly in spheroplasts. It has also allowed us to propose protein identities (and eliminate others from consideration) for many of the ions observed in MALDI (and ESI-MS) whole cell profiling at a specified growth condition. (J Am Soc Mass Spectrom 2002, 13, 804–812) © 2002 American Society for Mass Spectrometry

Microbiological applications of matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS) have piqued the interest of mass spectrometrists. Whole cell-derived protein fingerprints (signatures) show promise in bacterial identification/differentiation for clinical, environmental, and defense applications, and streamline analyses of proteins over-expressed in bacteria [1–33]. A recognized limitation of bacterial fingerprinting and other methods that supply only intact masses, however, has been their inability to *unambiguously* identify the proteins or other biomolecules responsible for the observed spectra [4, 18, 22, 34–39]. The confidence ascribed to identifications increases significantly for a well-characterized bacterium such as *Escherichia coli*, because considerable information regarding post-translational modifications

has been uncovered, and is readily accessed via carefully annotated databases such as Swiss-Prot [40]. Moreover, predictions of open reading frames, cleaved signal peptides [41, 42], and excised initiator methionines [43–46], are also reasonably accurate, although not infallible. This knowledge base and higher accuracy mass measurements e.g., via Fourier transform mass spectrometry (FTMS), increase the degree of confidence further, but additional search constraints, such as amino acid sequence, remain highly desirable. In this study, we explore the potential of metabolic labeling [47–53] to provide additional constraints linking protein signatures to protein identities in whole cell fingerprinting experiments, and exploit these capabilities to deliver insights into an antibacterial drug's mechanism of action.

Experimental

MALDI-MS employed a PerSeptive Voyager Elite time-of-flight mass spectrometer (Applied Biosystems, Framingham, MA) equipped with delayed extraction and

Published online May 15, 2002

Address reprint requests to Dr. R. Ogorzalek Loo, Department of Biological Chemistry, UCLA, 406A Boyer Hall, Los Angeles, CA 90095-1570, USA.
E-mail: rloo@mednet.ucla.edu

*Also at the Department of Chemistry and Biochemistry, UCLA.

operated in linear mode. Sinapinic acid (*trans*-3,5-dimethoxy-4-hydroxycinnamic acid), ferulic acid (*trans*-4-hydroxy-3-methoxycinnamic acid), and gentisic acid (2,5-dihydroxybenzoic acid; DHB) were employed as matrices. Sinapinic acid was prepared as a saturated solution in either 33% CH₃CN/67% H₂O/0.1% trifluoroacetic acid (TFA) (vol/vol/vol) or 50% CH₃CN/50% H₂O/0.1% TFA. DHB was prepared as an approximately 20 mg/mL solution in 50% CH₃CN/50% H₂O/0.1% TFA. Ferulic acid (saturated) was also prepared in 50% CH₃CN/50% H₂O/0.1% TFA. Matrices were purchased from Aldrich (Milwaukee, WI) or Sigma (St. Louis, MO). Anhydrous methanol (AR-ACS grade) and HPLC-grade water and acetonitrile were acquired from Mallinckrodt Baker, Inc. (Paris, KY), while sequencing grade TFA was acquired from Perkin Elmer–Applied Biosystems Division (Warrington, Great Britain).

Calibration of the time-of-flight MS was performed either externally (two-point calibration with myoglobin and insulin singly charged ions) or internally (single point calibration assuming a mass of 7272 for peaks in the 7270–7275 Da region). Repeated analyses on our Voyager Elite verified that a 1 point calibration in the 7100–7300 Da region (relying heavily on instrument default parameters) yielded better than $\pm 0.1\%$ mass accuracy below 30 kDa (typically $<0.05\%$ below 20 kDa). Sinapinic acid was preferred as a MALDI matrix for these analyses due to its consistently good mass accuracy, resolution, and sensitivity. Occasionally DHB or ferulic acid was also employed, because of their apparently higher tolerance to non-protein impurities. Relative peak intensities varied somewhat with different matrices; the most striking difference was that of the 15,879 Da peak's intensity, as described later.

LC-ESI-MS analysis was performed on an ion trap mass spectrometer (LCQ, Finnigan MAT, San Jose, CA) coupled on-line to a capillary HPLC (MicroPro, Eldex Laboratories, Napa, CA) and a microautosampler (Alcott 718 AL, Norcross, GA) in conjunction with a protein trapping pre-column (CapTrap, Michrom BioResources, Auburn, Ca). The samples were pre-concentrated and desalted on the CapTrap prior to eluting onto the capillary column (0.3 \times 150 mm PepMap C-18 or C-8, 5 μ m, 300 Å; LC Packings, San Francisco, CA). A step gradient was used to deliver solvents from 20% B to 65% B during the initial 20 min, then increased to 75% B for 5 min, and finally ramped up to 95% B for the final 5 min (30 min total gradient time). Mobile phase A consisted of 0.1% formic acid, 2% acetonitrile, 98% water and mobile phase B contained 0.1% formic acid, 20% water, and 80% acetonitrile. The HPLC was run at a flow rate of 5 μ L/min.

An overnight culture was prepared by inoculating 20 mL of Luria broth (Gibco) with *E. coli* EP1581, a K-12 type strain derived from W3110 [54, 55], possessing diminished drug efflux capability. The culture was grown aerobically, at 37 °C, with 200 RPM agitation. An aliquot was harvested, resuspended in 130 mL glucose-limiting potassium morpholinopropane sulfonate

(MOPS) medium [56], and incubated as above. Upon reaching an OD₆₀₀ of 0.14, 5.0 mL aliquots of the culture were deposited in 20 mL scintillation vials containing L-methionine or D,L-selenomethionine (Sigma S3875, St. Louis, MO), achieving 200 μ g/mL Met or Met-analogue concentrations [57]. At this time, compounds of interest solubilized in dimethyl sulfoxide (DMSO) or DMSO controls were dispensed to selected vials, yielding drug concentrations of 0.5–50 μ g/mL and 0.1–0.5% DMSO. Vials were incubated aerobically for 2.5 h at 37 °C with 200 RPM shaking. Cell growth was monitored periodically and immediately prior to harvest by withdrawing 0.5 mL of the culture, diluting 5-fold in glucose-limiting MOPS medium containing 0.2% formaldehyde, and measuring the absorbance of 600 nm light. (Absorbances were multiplied by 5 to yield the culture's optical density at each time point.) Cells were harvested by dispensing 1–1.5 mL solution to microcentrifuge tubes held on ice, centrifuging for 5 min at 15,000 RPM, 4 °C, and discarding the supernatant. Cell pellets were stored at –80 °C until use.

Prior to MALDI mass spectrometric analysis, water was added to each *E. coli* pellet according to the ratio: 40 μ L per A₆₀₀ unit per mL culture. (We estimate that 1 mL at A₆₀₀ = 1.0 corresponds to 200 μ g of total protein.) Following vortexing, 1 μ L of the *E. coli* suspension was added to 9 μ L of matrix solution, further vortexed, and 0.5 μ L was spotted to the sample probe and allowed to dry at room temperature. The *E. coli*/water suspension and matrix/sample mixtures were stored at –80 °C for subsequent measurements.

For LC-MS analysis, cell suspensions in water were subjected to three or more freeze thaw cycles and pelleted. The supernatant was removed, and the protein therein precipitated in 500 μ L of 10% trichloroacetic acid/90% acetone/20 mM dithiothreitol (DTT) at –20 °C. The precipitate was washed with cold acetone/20 mM DTT and dried. Prior to LC-MS the sample was dissolved in 100 μ L of 0.09% TFA/10% CH₃CN/90% H₂O.

Results and Discussion

Among the most important and difficult questions in drug discovery is that of a compound's mechanism of action or toxicity. No defined procedure exists to divine the answer, but global analytical methods such as two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) and DNA microarrays are valuable tools employed in its pursuit. Because these methods have limitations, e.g., analysis of basic proteins and post-translational modifications, genetic approaches also figure prominently, i.e., mapping resistance mutations. Despite these state-of-the art tools, answers remain elusive. Biological responses are complex and linked at many levels. Even when gene products are observed to change in abundance, the biological relevance may lie elsewhere, with primary and secondary responses assembled more subtly than anticipated. It is because of

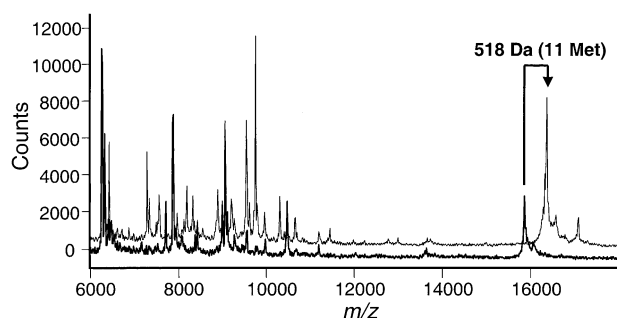


Figure 1. MALDI mass spectra of *E. coli* dosed with a stressor and cultured in methionine-supplemented media (bottom spectrum) and dosed with a stressor, but cultured in selenomethionine-supplemented media (top spectrum).

this complexity that alternative analytical methods capable of addressing these issues and/or complementing current capabilities are sought.

Protein Identification

MALDI-MS whole cell profiling [26] is one of these complementary approaches. Its speed and simplicity have accelerated its acceptance in microbiological applications, but ambiguities in protein identification restrain its potential. Efforts addressing protein identities have included correlating a m/z 9743 peak to the presence of the *E. coli* F-plasmid [36] and isolating *E. coli* proteins *hdeA* and *hdeB* to demonstrate their potential to yield peaks at m/z 9742 and 9066 [38]. With a similar HPLC-based strategy, Dai et al. [18] argued that peaks at 7272 and 7333 m/z arose from *cspC* and *cspE*, respectively, while the preparative isoelectric focusing (IEF) strategy of Arnold et al. [4] led them to ascribe m/z 7334 to entry exclusion protein b. Subsequently, mass measurements of intact ribosomes [5, 58] demonstrated the likely contributions of ribosomal proteins to MALDI profiles, including *r129* at m/z 7274, and the likelihood that abundant, basic proteins engender characteristic peaks [22]. Most notable from these results, is that multiple species *may* be responsible for a given peak in a bacterial profile, and that protein isolation strategies, while addressing a species' abundance and mass, cannot unambiguously assign it as the species responsible for a peak in a spectrum. Also not to be overlooked, is that results obtained under one set of conditions (e.g., bacterial strain, environment, sample preparation) do not necessarily transfer to a different set of conditions.

MALDI-MS spectra were obtained from *E. coli* cells dosed with an antibacterial agent of interest. We noted the striking appearance of an intense peak at m/z 15,879, especially dominant when ferulic acid was employed as the MALDI matrix (Figure 1). Its prominence increased with increasing doses of antibacterial agent. Hoping to relate the induced protein to the drug's mechanism, we sought its identity. Searching the SWISS-PROT database's *E. coli* entries by intact mass ($\pm 0.1\%$), revealed five contenders: (1) *f6b2*, CS6 fimbrial subunit B precursor,

pI 4.55, M_r 15,877; (2) *flgN*, flagella synthesis protein, pI 6.41, M_r 15,867, (3) *kla2*, antirestriction protein, pI 5.22, M_r 15,867; (4) *pi00*, pino protein, pI 4.78, M_r 15,876; and (5) *spY*, spheroplast protein Y, pI 9.45, M_r 15,876. The *spY* mass reflects its mature form, lacking the signal peptide. A search was also performed for a mass 131 Da larger, to account for proteins trimmed of their initiator methionine. (SWISS-PROT entries reflect methionine excision only when it has been established; e.g., by Edman degradation.) The second search added *rl13*, *trf5*, and *up03* as matches, but both ribosomal protein L13 and *up03* were immediately discounted; they are known to retain their initiator methionines, as reflected in database annotations. Because we employed an *E. coli* strain free of plasmids, the plasmid proteins *kla2* and *trf5* were also discarded. That many plasmid-encoded proteins are small, in concert with bacterial fingerprinting's superior detection of lower molecular weight proteins (i.e., <20 kDa), leads to a seemingly inordinate number of plasmid proteins arising as potential identifications. Clearly, analytical complexity is reduced when one is favored with knowledge as to the presence or absence of particular plasmids. Similarly, all-encompassing databases contain a number of strain-specific entries. By this logic, we could eliminate protein *f6b2* from consideration; it was not sequenced in our strain.

The protein *flgN* ought to be eliminated based on the 0.03% mass accuracy observed (validated by other proteins in these spectra), but choosing between *pi00* (M_r 15876 Da) and *spY* (M_r 15876 Da) is more difficult. One can posit the likelihood of observing *spY*, based on its higher isoelectric point [22], but such an argument would be far from definitive. The possibility that *both* proteins could contribute should not be neglected, nor should the possibility of heretofore undetected modifications, processing events, missed open-reading frames, or alternative initiation sites.

Seeking to differentiate between possibilities while assuming little, we considered *in vivo* labeling. Stable isotope labeling has provided constraints useful to both peptide [50, 51, 59] and protein mass analysis [53]. Analogous efforts securing additional constraints for protein identification [60] have included cysteine alkylation [52, 61], H/D exchange [62, 63], acylation [63], and other derivatizations [64, 65].

For our application, the ideal approach would yield a dramatic shift in m/z (demanding little of mass accuracy and resolution) and be interpreted easily. Based on these parameters, we considered culturing bacteria with selenomethionine. The large 46.89 Da mass difference between selenium and sulfur should display the proteins' methionine compositions readily. That the amino acid is relatively rare, comprising 1.7% of a protein's amino acids on average, simplifies the analysis [66]. Also invaluable is the considerable experience available in methionine-labeled proteins: ^{35}S -Met radiolabeling is a fundamental technology in proteomics and Se-Met labeling has found utility in protein crystallography.

Table 1. Proteins potentially yielding m/z 15,879

Protein	Average molecular weight	Isoelectric point (pI)	# Methionines expected
<i>f6b2</i>	15,877	4.55	3
<i>flgN</i>	15,867	6.41	6
<i>pi00</i>	15,876	4.78	2
<i>spY</i>	15,876	9.45	11

Selenomethionine- ^{75}Se has been injected intravenously for studies of human lipoprotein metabolism [67], while dual-label autoradiography (^{35}S - and ^{75}Se -methionine) facilitated discovery of proteins specific to disease states [68]. Methionine labeling of *E. coli* does not require use of an auxotrophic strain [53] (compromised in ability to synthesize an amino acid), a valued characteristic when other parameters must guide the choice of strain; e.g., antibiotic susceptibility or resistance. The only requirement of methionine labeling is ability to culture in chemically defined media, so that the pool of methionine available can be controlled.

Table 1 lists the number of methionines anticipated in the candidate proteins. Cells cultured and prepared as described previously yielded the spectra illustrated in Figure 1. Clearly many ions have shifted in m/z ; of most interest is the 518 Da shift at 16 kDa, indicative of a protein with 11 methionines ($518 \div 47 = 11$). This simple experiment eliminated all but one contender, and only alternatives with 11 Mets deserve future consideration.

More challenging is the question of whether the identification of a protein lends insight into biological activity. This mass spectrometry information will be combined with other assays and 2-D polyacrylamide gel electrophoretic analyses correlating spot patterns to antibiotics with known mechanisms of action to, hopefully, signal the important pathways involved in the cells' reaction to this compound. It has been postulated that spheroplast protein Y, a secreted protein not detected immunologically in intact cells, but produced abundantly in spheroplasts, is a stress protein for spheroplasting [69]. Thus, we speculate that the *E. coli* cells sense perturbations in their outer membrane effected by the compound; they respond by inducing *spY*.

Protein Synthesis under Dynamic Conditions

Organisms' responses to environmental changes or to various stresses beget dimensions of information beyond the simple catalogue of gene products of which they are molded. In applications targeting the rapid identification of microorganisms, for which reproducibility reigns, laboratories have sought to limit these dimensions, striving to produce images of a static proteome, one most amenable to inter-laboratory comparison [19, 24, 27]. But for other applications, viewing the dynamic landscape of cellular response is paramount, and in that venue, radioisotope labeling is

preferred. The proteomic mass spectrometry realm grasped stable isotope labeling to quantify relative protein levels, exploiting both in vitro and in vivo labeling [59, 61, 70, 71]. However, MS-based in vivo labeling methods have so far lacked pulse labeling's gift of distinguishing proteins synthesized prior to the stress from those synthesized later.

Timing the addition of selenomethionine to coincide with drug administration, we can differentiate proteins synthesized after dosing from those synthesized before. For example, the upper selenomethionine trace in Figure 1 shows no 15,879 Da (0% Se) peak; *spY* is present entirely as the 16,397 Da (100% Se) protein. Clearly, *spY* is induced upon addition of the drug and all of the 15.9 kDa protein in the lower, S-methionine trace, was synthesized after dosing. The advantage for quantification is clear: Shifting aside background proteins synthesized prior to application of the stressor provides results of higher precision, and enhances ability to detect subtle alterations. Utilizing this experimental design, we realize the large advantage that metabolic labeling enjoys over post-harvest labeling.

Suggested Changes and Variations for Identifying and Quantifying Proteomes

To quantify effects of stressors, we would choose to compare a control culture selectively labeled with amino acid analogue "A" at time t_1 , to a second culture, which at time t_1 , was also labeled with "A" and dosed with compound "X". Implementing labels in this manner would shift Met-containing peaks in both cultures, enabling comparisons of proteins synthesized over equivalent lengths of time, while embedding a proteome of normalization factors (unlabeled proteins synthesized prior to the stress). Care should be taken when interpreting ratios of peaks composed of more than one protein. Our protein-by-protein analysis for the compound inducing *spY* is in progress.

In identifying proteins for dynamic effects, the cleanest approach conceptually encompasses two sets of identifications—(1) proteome of unstressed cells, and (2) proteome of stressed cells. Set (1) reflects the approach taken in preceding work [49, 51, 53, 70]: Grow 1 set of cells in standard media and compare them to a second set of cells grown under the same conditions, except labeled with "A" throughout the culture. Identifications will rely on the mass and the constraint embedded in the number of "A" labels. Set (2) may employ single or double labeling. Single labeling was undertaken by us to identify *spY*: Application of a stress without amino acid labeling versus application of a stress with concomitant amino acid labeling. The only drawback with single-labeling is that for the unlabeled culture, it does not differentiate proteins synthesized before the stress from those synthesized after, potentially complicating the analysis. However, the complication may be unimportant, and would likely be illu-

minated by evaluating the quantification and unstressed proteome data, described above. Alternatively, double labeling could be pursued with two cultures: At time t_1 , culture 1 is selectively labeled with amino acid analogue "A" and dosed with compound "X", while culture 2 is labeled with amino acid analogue "B" and similarly dosed with compound "X". The shifts between B- versus A-labeled proteins deliver direct readouts of the number of methionines in proteins synthesized after the stressor. The main drawbacks of double-labeling are its potential to complicate the spectra and the need to invest in a second label or analogue.

While the replacement of methionine by selenomethionine worked well in this application, we would generally recommend implementation of stable-isotope labeled methionine for antibacterial studies. Selenomethionine lengthens the generation time of *E. coli* 26, and was reported to markedly inhibit growth of *E. coli* K12 [72]. Our observations were consistent; *E. coli* W3110 grew slower in the selenomethionine-supplemented media than in the methionine media. The selenomethionine bacteria also appeared to be inhibited less by the antibacterial compounds examined, which we believe can be attributed to their slower growth. Fortunately several forms of stable-isotope labeled methionine are available, reflecting its utility in protein NMR studies. While they do not shift protein masses as generously, the 13 Da shift of $^2\text{H}_8\text{-}^{13}\text{C}_5$ -methionine is significant. Methionine labeling provides considerable flexibility for double-label experiments, through the availability of deuterated, ^{13}C -labeled, and multiply labeled analogues, as well as the seleno-analogue investigated by us.

Labeling of other amino acids should be considered to complement methionine labeling. For example, proteins synthesized after addition of the label, but possessing no methionines, cannot be differentiated from proteins synthesized before addition of the label. Incorporation of a different analogue will differentiate synthesis before and after an applied stress for another set of proteins, as well as provide additional constraints for protein identification. Methionine labeling is attractive because it does not rely on auxotrophic strains, e.g., the presence of 0.1 mM D,L-selenomethionine reduces methionine biosynthesis by 95% [72]. Insight into amino acid interconversion and labeling strategy gleaned from previous radiolabeling efforts suggests other labeling conditions suitable for non-auxotrophic strains [47, 48, 73].

Identifying the Protein Ions Observed in Whole-Cell Profiles

Each ion mass was searched against the SWISS-PROT database at 0.05% mass accuracy using TagIdent [74], with the broadest range of isoelectric points available. Considerations included: (1) Correcting polypeptide

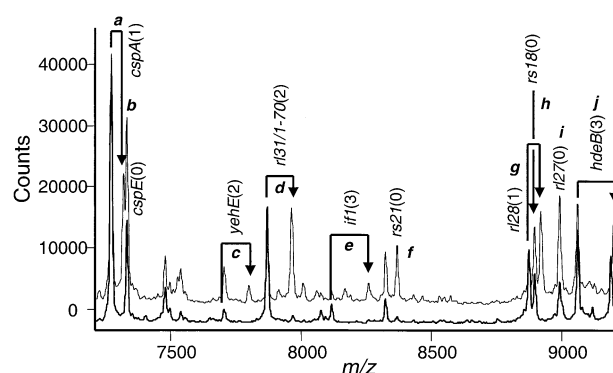


Figure 2. MALDI mass spectra of unstressed *E. coli* cultured in methionine-supplemented media (bottom spectrum) and cultured in selenomethionine-supplemented media (top spectrum). Numbers in parentheses denote the number of methionine residues in the proteins. Proteins consistent in mass, but eliminated by number of methionines include: Region *a*-*rl29*(2), region *c*-*yahO*(4), region *e*-*yqgC*(2), region *f*-*yqfE*(3) and *feoA*(2), region *h*-*ydhI*(3), *yahM*(3), and *yjbO*(5), region *i*-*yecJ*(2), and region *j*-*p76358*(1). Proteins consistent in mass that may also contribute to the spectra include region *a*-*cspC*(0) and region *g*-*yfdY*(7). Proteins with no methionines can rarely be eliminated from consideration, due to mass overlap with protein synthesized before labeling. The strength of the 7273 Da peak in the selenomethionine spectrum, however, argues that *cspC* is a significant presence. This data cannot dismiss potential contributions to *g* from *yfdY* because a peak is observed at the seven methionine increment. That peak is also consistent with 3-selenium shifted *hdeB*.

masses for detachment of signal peptides (The TagIdent utility performed this subtraction.); (2) accounting for post-translational modifications known or suspected; (3) adding 787 Da to the polypeptide minus signal peptide mass of lipid-acylated proteins. The lipid increment was based on previous studies of *E. coli* lipoproteins *mull* and *osmE* [39]; (4) discarding plasmid proteins and proteins undocumented in the K12 substrains MG1655 or W3110; (5) compensating for undocumented initiator methionine excision by searching for masses incremented by 131 Da.

For all sequences with uncharacterized amino termini, the likelihood of initiator methionine excision was evaluated based on the neighbor amino acid [43–46, 75, 76]: Met-(Ala, Cys, Gly, Pro, Ser, and Thr) were classified as likely to cleave. Met-(Asp, Asn, Leu, Ile, Val, Gln, Glu, His, Met, Phe, Lys, Tyr, Trp, and Arg) were classified as unlikely to cleave. Although Hirel et al. [46] list Met-Val as likely to be cleaved, no examples as such were observed in the Edman degradation results of Link [75].

Figures 2, 3, and 4 illustrate spectra obtained from *E. coli* drug-free cultures, labeled with either S-Met or Se-Met, as described previously.

We find it interesting that the peak at 7273 arises from *cspA* and *cspC* (Figure 2), but not *rl29*, despite the prevalence of other peaks attributed to ribosomal proteins. Proteins of the cold-shock domain family, to which *cspA* and *cspC* belong, are induced by various stresses. We expect their abundances and relative contributions to vary dramatically with growth conditions,

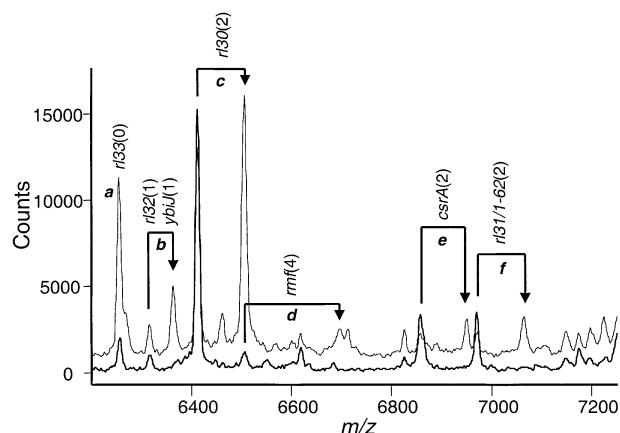


Figure 3. MALDI mass spectra of *E. coli* cultured in methionine-supplemented media (bottom spectrum) and cultured in selenomethionine-supplemented media (top spectrum). Numbers in parentheses denote the number of methionine residues in the proteins. Note that *rl32* and *yblJ* both have one methionine; either or both proteins could give rise to the region *b* peaks. A region *e* protein consistent in mass, but eliminated by its number of methionines is *ycaR*(1).

factors which should be considered if the *m/z* 7273 ion is employed as a marker. Also of note was that under these culture conditions, the *m/z* 13,002 ion arose not from the expected ribosomal protein *rl19*, but primarily from the hypothetical protein *yifE*. Clearly, protein identities need to be verified for each growth condition of interest.

Despite the surprises described above, most of the selenium-related shifts observed were consistent for ribosomal proteins at their expected *m/z* values (19 ribosomal proteins confirmed). In addition, a protein at

m/z 6317 can be ascribed to either *rl32* or mature *yblJ* (detached signal peptide); both potential proteins contain a single methionine. Non-ribosomal proteins matching in *m/z* and number of methionines included: *rmf* (6507 Da), *csrA* (6858 Da), *cspC* and *cspA* (7273 Da), *cspE* (7333 Da), *yehE* (7707 Da), *cspD* (7970 Da), *if1-Met* (8119 Da), *hdeB* (9064 Da), *dbhB* (9226 Da), *dbhA* or *tatA-Met* (9536 Da), *hdeA* (9740 Da), *ygiW* (11,796 Da), *yifE* (13,002 Da), and *osmY* (18,160 Da).

The identities of several ions are unresolved. An ion observed at 5065 Da did not display the two methionine shift, anticipated for *ydaG-Met* (expected 5064 Da), the only possibility within mass tolerances, but of low probability because *ydaG* was expected to retain its initiator Met. No protein matched within 0.05% mass tolerance to a 5459 Da ion found to contain one methionine. An 8326 *m/z* ion did not contain the two methionines required for assignment to *yjbJ* or *ydaQ-Met* (both ions predicted at 8326 *m/z*), instead, it contained zero or one methionine (overlap with another protein prevented us from differentiating between the two possibilities). Ions observed at 10,397 and 10,797 *m/z* did not contain the two methionines predicted for *yjbD-Met* and *yhcO*, respectively, instead the former possessed none, while the latter contained either 0 or 1 (ambiguity introduced by overlap with another protein). Finally, an ion observed at 10,652 Da contained three methionines, consistent with assignment as *ihfB* (predicted *m/z* 10,652), but we are less confident, because initiator methionine excision is highly probable for this protein's Met-Thr amino terminus.

An alternative method for profiling intact proteins employs LC-MS [12]. Figure 5 illustrates a selenomethionine versus sulfur-methionine comparison performed by LC-MS with electrospray ionization. Although the LC separation was less than optimal for such a complex mixture, the concept is demonstrated by this example. On-line LC-MS/MS of the 13.6 kDa protein (top, Figure 5) further confirmed the identity of the *rs12* protein (data not shown). The "top-down" approach for protein identification, in which molecular mass measurement of the intact protein combined with sequence information from fragmentation of the gas phase protein molecule can proceed by in-source decay of MALDI-generated ions [39] or MS/MS of ESI-generated ions [77, 78].

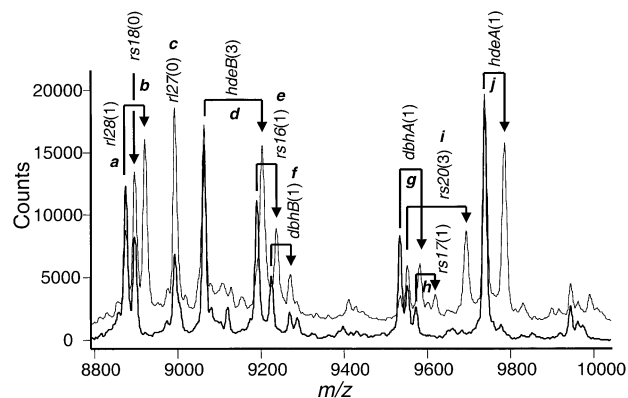


Figure 4. MALDI mass spectra of *E. coli* cultured in methionine-supplemented media (bottom spectrum) and cultured in selenomethionine-supplemented media (top spectrum). Numbers in parentheses denote the number of methionine residues in the proteins. Note that *dbhA*, *tatA-Met*, and *yceP-Met* all have one methionine; although only *dbhA* is labeled in the spectrum, any combination of these proteins generate the region *g* peaks. Proteins consistent in mass, but eliminated by number of methionines include: Region *b*—*ydhI*(3), *yahM*(3), and *yjbO*(5), region *c*—*yecJ*(2), region *d*—*p76358*(1), region *e*—*yfgJ*(4), and region *h*—*ilvM*(3). As discussed in the Figure 2 caption, this data cannot dismiss potential contributions from *yfdY*(7).

Conclusions

These data illustrate a method providing additional information to support or refute protein identities in MALDI-MS of whole cells. It does not require the use of auxotrophic strains. More powerful is that it allows one to selectively label proteins synthesized after a stress, e.g., a change in temperature, environment, or addition of a drug. Labeling in this manner, analogous to the methods of radiolabeling, provides quantitative information on the dynamics of the cellular proteome's response to perturbation.

23. Ryzhov, V.; Hathout, Y.; Fenselau, C. Rapid Characterization of Spores of *Bacillus Cereus* Group Bacteria by Matrix-Assisted Laser Desorption-Ionization Time-of-Flight Mass Spectrometry. *Appl. Environ. Microbiol.* **2000**, *66*, 3828–3834.
24. Saenz, A. J.; Petersen, C. E.; Valentine, N. B.; Gantt, S. L.; Jarman, K. H.; Kingsley, M. T.; Wahl, K. L. Reproducibility of Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry for Replicate Bacterial Culture Analysis. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1580–1585.
25. Gantt, S. L.; Valentine, N. B.; Saenz, A. J.; Kingsley, M. T.; Wahl, K. L. Use of an Internal Control for Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry Analysis of Bacteria. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1131–1137.
26. Wilcox, S. K.; Cavey, G. S.; Pearson, J. D. Single Ribosomal Protein Mutations in Antibiotic-Resistant Bacteria Analyzed by Mass Spectrometry. *Antimicrob. Agents Chemother.* **2001**, *45*, 3046–3055.
27. Wang, Z.; Russon, L.; Li, L.; Roser, D. C.; Long, S. R. Investigation of Spectral Reproducibility in Direct Analysis of Bacterial Proteins by MALDI-TOFMS. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 456–464.
28. Haag, A. M.; Taylor, S. N.; Johnston, K. H.; Cole, R. B. Rapid Identification and Speciation of *Haemophilus* Bacteria by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *J. Mass Spectrom.* **1998**, *33*, 750–756.
29. Lynn, E. C.; Chung, M.-C.; Tsai, W.-C.; Han, C.-C. Identification of *Enterobacteriaceae* Bacteria by Direct Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis of Whole Cells. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2022–2027.
30. Leenders, F.; Stein, T. H.; Kablitz, B.; Franke, P.; Vater, J. Rapid Typing of *Bacillus subtilis* Strains by Their Secondary Metabolites Using Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Intact Cells. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 943–949.
31. Nilsson, C. L. Fingerprinting of *Helicobacter pylori* Strains by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometric Analysis. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 1067–1071.
32. Edwards-Jones, V.; Claydon, M. A.; Evason, D. J.; Walker, J.; Fox, A. J.; Gordon, D. B. Rapid Discrimination Between Methicillin-Sensitive and Methicillin-Resistant *Staphylococcus aureus* by Intact Cell Mass Spectrometry. *J. Med. Microbiol.* **2000**, *49*, 295–300.
33. Winkler, M. A.; Uher, J.; Cepa, S. Direct Analysis and Identification of *Helicobacter* and *Campylobacter* Species by MALDI-TOF Mass Spectrometry. *Anal. Chem.* **1999**, *71*, 3416–3419.
34. Cavalcoli, J. D.; VanBogelen, R. A.; Andrews, P. C.; Moldover, B. Unique Identification of Proteins from Small Genome Organisms: Theoretical Feasibility of High Throughput Proteome Analysis. *Electrophoresis* **1997**, *18*, 2703–2708.
35. Anderson, G.; Bruce, J. E.; Pasa Tolic, L.; Smith, R. D. *Proceedings of the 46th Conference of the American Society for Mass Spectrometry*; Orlando, FL, 1998, p 1270.
36. Karty, J. A.; Lato, S.; Reilly, J. P. Detection of the Bacteriological Sex Factor in *E. coli* by Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 625–629.
37. Arnold, R. J.; Reilly, J. P. Observation of Tetrahydrofolypolyglutamic Acid in Bacteria Cells by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Anal. Biochem.* **2000**, *281*, 45–54.
38. Holland, R. D.; Duffy, C. R.; Rafii, F.; Sutherland, J. B.; Heinze, T. M.; Holder, C. L.; Voorhees, K. J.; Jackson, O. L., Jr. Identification of Bacterial Proteins Observed in MALDI TOF Mass Spectra from Whole Cells. *Anal. Chem.* **1999**, *71*, 3226–3230.
39. Loo, R. R.; Cavalcoli, J. D.; VanBogelen, R. A.; Mitchell, C.; Loo, J. A.; Moldover, B.; Andrews, P. C. Virtual 2-D Gel Electrophoresis: Visualization and Analysis of the *E. coli* Proteome by Mass Spectrometry. *Anal. Chem.* **2001**, *73*, 4063–4070.
40. Bairoch, A. R.; Apweiler, R. A. The SWISS-PROT Protein Sequence Database and Its Supplement TrEMBL in 2000. *Nucleic Acids Res.* **2000**, *28*, 45–48.
41. Edman, M.; Jarhede, T.; Sjöström, M.; Wieslander, A. Different Sequence Patterns in Signal Peptides from Mycoplasmas, Other Gram-Positive Bacteria, and *Escherichia coli*: A Multivariate Data Analysis. *Prot. Struct. Funct. Genet.* **1999**, *35*, 195–205.
42. von Heijne, G. A New Method for Predicting Signal Sequence Cleavage Sites. *Nucleic Acids Res.* **1986**, *14*, 4683–4690.
43. Ben-Bassat, A.; Bauer, K.; Chang, S. Y.; Myambo, K.; Boosman, A.; Chang, S. Processing of the Initiation Methionine from Proteins: Properties of the *Escherichia coli* Methionine Aminopeptidase and Its Gene Structure. *J. Bacteriol.* **1987**, *169*, 751–757.
44. Boissel, J. P.; Kasper, T. J.; Shah, S. C.; Malone, J. I.; Bunn, H. F. Amino-Terminal Processing of Proteins: Hemoglobin South Florida, a Variant with Retention of Initiator Methionine and N^α-Acetylation. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 8448–8452.
45. Dalboge, H. S.; Bayne, S.; Pedersen, J. In Vivo Processing of N-Terminal Methionine in *Escherichia coli*. *FEBS Lett.* **1990**, *266*, 1–3.
46. Hirel, P.-H.; Schmitter, J.-M.; Dessen, P.; Fyatt, G.; Blanquet, S. Extent of N-Terminal Methionine Excision from *Escherichia coli* Proteins Is Governed by the Side Chain Length of the Penultimate Amino Acid. *Biochemistry* **1989**, *28*, 8247–8251.
47. Garrels, J. I.; Futcher, B.; Kobayashi, R.; Latter, G. I.; Schwender, B.; Volpe, T.; Warner, J. R.; McLaughlin, C. S. Protein Identifications for a *Saccharomyces cerevisiae* Protein Database. *Electrophoresis* **1994**, *15*, 1466–1486.
48. Maillet, I.; Lagniel, G.; Perrot, M.; Boucherie, H.; Labarre, J. Rapid Identification of Yeast Proteins on Two-Dimensional Gels. *J. Biol. Chem.* **1996**, *271*, 10263–10270.
49. Chen, X.; Fei, Z.; Smith, L. M.; Bradbury, E. M.; Majidi, V. Stable-Isotope-Assisted MALDI-TOF Mass Spectrometry for Accurate Determination of Nucleotide Compositions of PCR Products. *Anal. Chem.* **1999**, *71*, 3118–3125.
50. Chen, X.; Smith, L. M.; Bradbury, E. M. Site-Specific Mass Tagging with Stable Isotopes in Proteins for Accurate and Efficient Protein Identification. *Anal. Chem.* **2000**, *72*, 1134–1143.
51. Hunter, T. C.; Yang, L.; Zhu, H.; Majidi, V.; Bradbury, E. M.; Chen, X. Peptide Mass Mapping Constrained with Stable Isotope-Tagged Peptides for Identification of Protein Mixtures. *Anal. Chem.* **2001**, *73*, 4891–4902.
52. Goodlett, D. R.; Bruce, J. E.; Anderson, G. A.; Rist, B.; Pasa-Tolic, L.; Fiehn, O.; Smith, R. D.; Aebersold, R. Protein Identification with a Single Accurate Mass of a Cysteine-Containing Peptide and Constrained Database Searching. *Anal. Chem.* **2000**, *72*, 1112–1118.
53. Veenstra, T. D.; Martinovic, S.; Anderson, G. A.; Pasa-Tolic, L.; Smith, R. D. Proteome Analysis Using Selective Incorporation of Isotopically Labeled Amino Acids. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 78–82.
54. Blattner, F. R.; Plunkett, G., III; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **1997**, *277*, 1453–1474.

55. Itoh, T.; Okayama, T.; Hashimoto, H.; Takeda, J.-I.; Davis, R. W.; Mori, H.; Gojobori, T. A Low Rate of Nucleotide Changes in *Escherichia coli* K-12 Estimated from a Comparison of the Genome Sequences between Two Different Substrains. *FEBS Lett.* **1999**, *450*, 72–76.
56. VanBogelen, R. A.; Neidhardt, F. C. Preparation of *Escherichia coli* Samples for Two-Dimensional Gel Analysis. Link, A., Ed. *In Methods in Molecular Biology: 2-D Proteome Analysis Protocols*, Vol. CXII. Humana Press: Totowa, 1998; pp 21–29.
57. VanBogelen, R. A. Generating a Bacterial Genome Inventory. Link, A., Ed. *In Methods in Molecular Biology: 2-D Proteome Analysis Protocols*, Vol. CXII. Humana Press: Totowa, 1998; pp 423–429.
58. Benjamin, D. R.; Robinson, C. V.; Hendrick, J. P.; Hartl, F. U.; Dobson, C. M. Mass Spectrometry of Ribosomes and Ribosomal Subunits. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7391–7395.
59. Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative Analysis of Complex Protein Mixtures Using Isotope-Coded Affinity Tags. *Nat. Biotechnol.* **1999**, *17*, 994–999.
60. Jensen, O. N.; Larsen, M. R.; Roepstorff, P. Mass Spectrometric Identification and Microcharacterization of Proteins from Electrophoretic Gels: Strategies and Applications. *Prot. Struct. Funct. Genet. Suppl.* **1998**, *2*, 74–89.
61. Sechi, S.; Chait, B. T. Modification of Cysteine Residues by Alkylation. A Tool in Peptide Mapping and Protein Identification. *Anal. Chem.* **1998**, *70*, 5150–5158.
62. James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. Protein Identification in DNA Databases by Peptide Mass Fingerprinting. *Protein Science* **1994**, *3*, 1347–1350.
63. Chaurand, P.; Luetzenkirchen, F.; Spengler, B. Peptide and Protein Identification by Matrix-Assisted Laser Desorption Ionization (MALDI and MALDI-Post-Source Decay Time-of-Flight Mass Spectrometry). *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 91–103.
64. Hale, J. E.; Butler, J. P.; Knierman, M. D.; Becker, G. W. Increased Sensitivity of Tryptic Peptide Detection by MALDI-TOF Mass Spectrometry Is Achieved by Conversion of Lysine to Homoarginine. *Anal. Biochem.* **2000**, *287*, 110–117.
65. Beardsley, R. L.; Karty, J. A.; Reilly, J. P. Enhancing the Intensities of Lysine-Terminated Tryptic Peptide Ions in Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 2147–2153.
66. Dayhoff, M. O. *Atlas of Protein Sequence and Structure*, Suppl. 3; National BioMedical Research Foundation: Washington, 1976.
67. Eaton, R. P.; Pommer, I. Incorporation of Selenium-75-Labeled Selenomethionine into Human Apoproteins. I. Characterization of Specificity in Very-Low-Density and Low-Density Lipoproteins. *Diabetes* **1976**, *25*, 32–43.
68. Easty, D. J.; Patel, K.; Dover, R.; Evans, D. J.; Dunn, M. J. A Comparative Study of Protein Synthesis by Keratinocytes and Fibroblasts in Vitro Using Two-Dimensional Gel Electrophoresis and Dual Isotope Autoradiography. *Electrophoresis (Weinheim, Germany)* **1988**, *9*, 227–231.
69. Hagenmaier, S.; Stierhof, Y.-D.; Henning, U. A New Periplasmic Protein of *Escherichia coli* Which Is Synthesized in Spheroplasts but Not in Intact Cells. *J. Bacteriol.* **1997**, *179*, 2073–2076.
70. Pasa-Tolic, L.; Jensen, P. K.; Anderson, G. A.; Lipton, M. S.; Peden, K. K.; Martinovic, S.; Tolic, N.; Bruce, J. E.; Smith, R. D. High Throughput Proteome-Wide Precision Measurements of Protein Expression Using Mass Spectrometry. *J. Am. Chem. Soc.* **1999**, *121*, 7949–7950.
71. Griffin, T. J.; Gygi, S. P.; Rist, B.; Aebersold, R.; Loboda, A.; Jilkine, A.; Ens, W.; Standing, K. G. Quantitative Proteomic Analysis Using a MALDI Quadrupole Time-of-Flight Mass Spectrometer. *Anal. Chem.* **2001**, *73*, 978–986.
72. Coch, E. H.; Greene, R. C. Utilization of Selenomethionine by *Escherichia coli*. *Biochim. Biophys. Acta* **1971**, *230*, 223–236.
73. Pedersen, S.; Bloch, P. L.; Reeh, S.; Neidhardt, F. C. Patterns of Protein Synthesis in *E. coli*: A Catalog of the Amount of 140 Individual Proteins at Different Growth Rates. *Cell* **1978**, *14*, 179–190.
74. Wilkins, M. R.; Gasteiger, E.; Bairoch, A.; Sanchez, J.-C.; Williams, K. L.; Appel, R. D.; Hochstrasser, D. F. Protein Identification and Analysis Tools in the ExPASy Server. *In 2-D Proteome Analysis Protocols*; Link, A. J., Ed.; Humana Press: Totowa, 1998; pp 531–552.
75. Link, A. J.; Robison, K.; Church, G. M. Comparing the Predicted and Observed Properties of Proteins Encoded in the Genome of *Escherichia coli* K-12. *Electrophoresis* **1997**, *18*, 1259–1313.
76. *Pfu Methionine Aminopeptidase*. Otsu, Shiga, Japan. Takara Biotechnology (Dalian) Co., Ltd., 2001.
77. Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J.-H.; Begley, T. P.; McLafferty, F. W. Top Down Characterization of Larger Proteins (45 kDa) by Electron Capture Dissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2002**, *124*, 672–678.
78. Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. Informatics and Multiplexing of Intact Protein Identification in Bacteria and the Archaea. *Nat. Biotechnol.* **2001**, *19*, 952–957.